УДК 004.655

А.С. Сенченко
Донбасский государственный педагогический университет, Украина
84116, Донецкая обл., г. Славянск, ул. Генерала Батюка, 19

Закон двойного отрицания и правило преобразования разности в табличных алгебрах

A.S. Senchenko
Donbas State Pedagogical University
Donetsk area, Slavyansk, st. Generala Batyuka, 19

A double Complement Law and a Difference Transformation Rule in Table Algebra

О.С. Сенченко
Донбаський державний педагогічний університет, Україна
84116, Донецька обл., м. Слов'янськ, вул. Генерала Батюка, 19

Закон подвійного заперечення та правило перетворення різниці у табличних алгебрах

В работе найдены необходимые и достаточные условия, при которых в табличных алгебрах выполняются закон двойного отрицания и правило преобразования разности. Приведены примеры, иллюстрирующие данные критерии.
Ключевые слова: табличная алгебра, активное дополнение, разность, база данных.

In this paper is found the necessary and sufficient conditions due to which in table algebra are fulfilled the double complement law and the difference transformation rule. This conditions are illustrated in exemples.
Key words: table algebra, active complement, difference, database.

У роботі знайдено необхідні та достатні умови, за яких у табличних алгебрах виконуються закон подвійного заперечення та правило перетворення різниці. Наведені приклади, що ілюструють знайдені критерії.
Ключові слова: таблична алгебра, активне доповнення, різниця, база даних.

Введение

В настоящее время системы управления базами данных широко используются во многих сферах человеческой деятельности. Наиболее распространённой является реляционная модель данных, впервые предложенная Э. Коддом в 1970 году [1]. С точки зрения математики реляционная база данных – конечный набор конечных отношений разной размерности между заранее определёнными множествами элементарных данных.

Табличные алгебры, введённые В.Н. Редько и Д.Б. Буем, построены на основе реляционных алгебр Э. Кодда и существенно их уточняют. Они составляют теоретический фундамент языков запросов современных табличных баз данных. Элементы носителя табличной алгебры уточняют реляционные структуры данных, а сигнатурные операции построены на базе основных табличных манипуляций в реляционных алгебрах и языке SQL.

В монографии [2] найдено и доказано большое количество различных свойств операций табличных алгебр. В настоящей работе найдены и доказаны необходимые и достаточные условия, при которых некоторые включения, доказанные в [2], превращаются в равенства.

Основные определения

Зафиксируем некоторое непустое множество $A = \{A_1, \dots, A_n\}$, элементы которого называются атрибутами. Произвольное конечное подмножество $R = \{A'_1, \dots, A'_k\} \subseteq A$ назовем схемой, причем схема может являться пустым множеством. Строкой s схемы R называется множество пар $s = \{(A'_1, d_1), \dots, (A'_k, d_k)\}$, проекция которого по первой компоненте равна R. Таблицей схемы R называется конечное множество строк схемы R. Далее в работе рассматриваем таблицы схемы R с количеством атрибутов k. На множестве всех таких таблиц введены операции объединения, пересечения и разности как аналогичные операции теории множеств; операции проекции, селекции, соединения (в некоторых источниках, например, в [3], эта операция называется эквисоединением), деления таблиц и операция переименования атрибутов; эти операции не будут использованы в настоящей работе, поэтому их определения не приводим. Также на множестве таблиц введена операция активного дополнения, для определения которой необходимо дать определение понятиям, используемым в работе в дальнейшем. Активным доменом атрибута A относительно таблицы T называется множество $D_{A,T} = \{d \mid \exists s \in T \wedge (A, d) \in s\}$, состоящее из всевозможных значений атрибута A в таблице T. Насыщением $C(T)$ называется таблица $\prod_{A \in R} D_{A,T}$, где $\prod$ – оператор прямого (декартового) произведения всех атрибутов схемы T. Другими словами, мы можем понимать насыщение как аналог декартового произведения активных доменов всех атрибутов таблицы в применении к именным множествам. Активным дополнением таблицы T называется таблица $\widetilde{T} = C(T) - T$.

Табличной алгеброй называют частичную алгебру с носителем – множеством всех таблиц произвольной схемы и приведёнными выше девятью операциями. В табличной алгебре выделяют две пустые таблицы: таблицу T_ε, схема которой является пустым множеством, и таблицу $T_\varnothing$ – пустое множество строк произвольной (в том числе и пустой) схемы.

Основные результаты

В монографии [2] в подразделе о свойствах насыщения и активного дополнения сформулирован и доказан ряд свойств этих операций, большая часть которых являются включениями. Автором были найдены необходимые и достаточные условия (в виде двух теорем), при которых этих включения превращаются в равенства для непустых таблиц, для пустых таблиц эти равенства тоже выполняются, но в этом случае могут не выполняться критерии.

Теорема 1 (закон двойного отрицания). При $T \neq T_\varnothing$ равносильны утверждения:

а) выполняется равенство $C(\widetilde{T}) = C(T)$;

б) выполняются равенства $\widetilde{\widetilde{T}} = T$;

в) для каждого значения x активного домена каждого атрибута A_q существуют такие значения $d_1, \dots, d_{q-1}, d_{q+1}, \dots, d_k$ активных доменов соответствующих атрибутов $A_1, \dots, A_{q-1}, A_{q+1}, \dots, A_k$ относительно таблицы T, что строка

$$\{(A_1, d_1), \dots, (A_{q-1}, d_{q-1}), (A_q, x), (A_{q+1}, d_{q+1}), \dots, (A_k, d_k)\} \notin T.$$

Доказательство. Покажем сначала эквивалентность утверждений (а) и (в). Пусть выполняется равенство $C(\tilde{T}) = C(T)$. По определению насыщения это равносильно выполнению равенств $\forall i\ D_{A_i,T} = D_{A_i,\tilde{T}}$. От противного, допустим, что для некоторого $x \in D_{A_q,T}$ для всех значений $d_1,\dots,d_{q-1},d_{q+1},\dots,d_k$ активных доменов остальных атрибутов $A_1,\dots,A_{q-1},A_{q+1},\dots,A_k$, все строки $s = \{(A_1,d_1),\dots,(A_{q-1},d_{q-1}),(A_q,x),(A_{q+1},d_{q+1}),\dots,(A_k,d_k)\} \in T$. Тогда, по определению активного дополнения, все эти строки не принадлежат таблице $\tilde{T}$, поэтому значение $x \notin D_{A_q,\tilde{T}}$, то есть $D_{A_q,T} \neq D_{A_q,\tilde{T}}$. Получившееся противоречие доказывает импликацию (а) $\Rightarrow$ (в).

Пусть теперь выполняется утверждение (в). От противного, допустим, что $C(\tilde{T}) \neq C(T)$. В монографии [2] (утверждение 2.2.1 пункт 4) доказано включение $C(\tilde{T}) \subseteq C(T)$, поэтому неравенство возможно только в том случае, когда существует такая строка $s' = \{(A_1,d'_1),\dots,(A_k,d'_k)\}$, что $s' \in C(T)$ и $s' \notin C(\tilde{T})$. Последнее выражение влечет существование некоторого индекса q, для которого $d'_q \notin D_{A_q,\tilde{T}}$. По предположению для некоторых значений $d_1,\dots,d_{q-1},d_{q+1},\dots,d_k$ активных доменов атрибутов $A_1,\dots,A_{q-1},A_{q+1},\dots,A_k$ относительно таблицы T строка $s = \{(A_1,d_1),\dots,(A_{q-1},d_{q-1}),(A_q,d'_q),(A_{q+1},d_{q+1}),\dots,(A_k,d_k)\} \notin T$, а поскольку $d'_q \in D_{A_q,T}$, то $s \in C(T)$, и значит $s \in \tilde{T}$, следовательно, $d'_q \in D_{A_q,\tilde{T}}$. Это противоречие доказывает импликацию (в) $\Rightarrow$ (а), то есть утверждения (а) и (в) равносильны.

Покажем теперь эквивалентность утверждений (а) и (б). Пусть выполняется равенство $\tilde{\tilde{T}} = T$. От противного, допустим, что $C(\tilde{T}) \neq C(T)$. По определению насыщения в этом случае не выполняется равенства $\forall i\ D_{A_i,T} = D_{A_i,\tilde{T}}$, следовательно, существует индекс q, для которого $D_{A_q,T} \neq D_{A_q,\tilde{T}}$. В монографии [2] (лемма 2.2.1 пункт 4) доказаны равенства $\forall i\ D_{A_i,C(T)} = D_{A_i,T}$, и поскольку по определению активного дополнения $\tilde{T} = C(T) - T$, то выполняется включение $D_{A_q,T} \supseteq D_{A_q,\tilde{T}}$. Следовательно, из неравенства $D_{A_q,T} \neq D_{A_q,\tilde{T}}$ и включения $D_{A_q,T} \supseteq D_{A_q,\tilde{T}}$ вытекает, что некоторое значение x входит в $D_{A_q,T}$ и не входит в $D_{A_q,\tilde{T}}$. Тогда по определению активного домена существует такая строка $s \in T$, что $(A_q,x) \in s$. Так как $x \notin D_{A_q,\tilde{T}}$, то по определению насыщения $x \notin C(\tilde{T})$, и следовательно $s \notin (C(\tilde{T}) - \tilde{T}) = \tilde{\tilde{T}}$, то есть $\tilde{\tilde{T}} \neq T$. Получили противоречие, что доказывает импликацию (б) $\Rightarrow$ (а).

Пусть теперь выполняется равенство $C(\tilde{T}) = C(T)$, что равносильно выполнению равенств $\forall i\ D_{A_i,T} = D_{A_i,\tilde{T}}$. От противного, допустим, что $\tilde{\tilde{T}} \neq T$. В [2] (утверждение 2.2.1 пункт 6) доказано включение $\tilde{\tilde{T}} \subseteq T$, поэтому неравенство $\tilde{\tilde{T}} \neq T$ возможно только тогда, когда существует такая строка s, что $s \in T$ и $s \notin \tilde{\tilde{T}}$. Пусть $s = \{(A_1,d_1),\dots,(A_k,d_k)\}$. Из $s \in T$, по свойствам насыщения следует, что $s \in C(T)$. Из

равенств $\forall i\; D_{A_i,T} = D_{A_i,\tilde{T}}$ следует, что $s \in C(\tilde{T})$. По определению активного дополнения $s \in T$ влечет $s \notin \tilde{T}$. Тогда получаем, что $s \in \tilde{\tilde{T}}$, что противоречит допущению $\tilde{\tilde{T}} \neq T$. Получившееся противоречие доказывает импликацию (а) $\Rightarrow$ (б), то есть утверждения (а) и (б) равносильны.

Проиллюстрируем критерий теоремы 1 на следующих примерах.

Пример 1.1.

$$\text{Пусть } T = \begin{array}{ccc} A & B & C \\ 1 & 2 & 2 \\ 1 & 3 & 1 \\ 2 & 2 & 2 \\ 2 & 2 & 3 \end{array}.$$

Значения активных доменов $D_{A,T} = \{1,2\}, D_{B,T} = \{2,3\}, D_{C,T} = \{1,2,3\}$. Для каждого значения активного домена каждого атрибута условие из (в) выполняется, равенства (а) и (б) должны выполняться. Действительно,

$$C(T) = \begin{array}{ccc} A & B & C \\ 1 & 2 & 1 \\ 1 & 2 & 2 \\ 1 & 2 & 3 \\ 1 & 3 & 1 \\ 1 & 3 & 2 \\ 1 & 3 & 3 \\ 2 & 2 & 1 \\ 2 & 2 & 2 \\ 2 & 2 & 3 \\ 2 & 3 & 1 \\ 2 & 3 & 2 \\ 2 & 3 & 3 \end{array},\quad \tilde{T} = \begin{array}{ccc} A & B & C \\ 1 & 2 & 1 \\ 1 & 2 & 3 \\ 1 & 3 & 2 \\ 1 & 3 & 3 \\ 2 & 2 & 1 \\ 2 & 3 & 1 \\ 2 & 3 & 2 \\ 2 & 3 & 3 \end{array},\quad C(\tilde{T}) = \begin{array}{ccc} A & B & C \\ 1 & 2 & 1 \\ 1 & 2 & 2 \\ 1 & 2 & 3 \\ 1 & 3 & 1 \\ 1 & 3 & 2 \\ 1 & 3 & 3 \\ 2 & 2 & 1 \\ 2 & 2 & 2 \\ 2 & 2 & 3 \\ 2 & 3 & 1 \\ 2 & 3 & 2 \\ 2 & 3 & 3 \end{array},\quad \tilde{\tilde{T}} = \begin{array}{ccc} A & B & C \\ 1 & 2 & 2 \\ 1 & 3 & 1 \\ 2 & 2 & 2 \\ 2 & 2 & 3 \end{array}.$$

$\tilde{\tilde{T}} = T$ и $C(\tilde{T}) = C(T)$.

Пример 1.2.

$$\text{Пусть } T = \begin{array}{ccc} A & B & C \\ 1 & 2 & 2 \\ 1 & 3 & 1 \\ 1 & 3 & 2 \\ 2 & 2 & 2 \\ 2 & 3 & 2 \end{array}.$$

Значения активных доменов $D_{A,T} = \{1,2\}, D_{B,T} = \{2,3\}, D_{C,T} = \{1,2\}$. Для значения $2 \in D_{C,T}$ все соответствующие строки в таблице T есть; условие из (в) не выполняется, равенства (а) и (б) не должны выполняться. Действительно,

$$C(T)=\begin{matrix} A & B & C\\ 1 & 2 & 1\\ 1 & 2 & 2\\ 1 & 3 & 1\\ 1 & 3 & 2\\ 2 & 2 & 1\\ 2 & 2 & 2\\ 2 & 3 & 1\\ 2 & 3 & 2 \end{matrix},\quad \widetilde{T}=\begin{matrix} A & B & C\\ 1 & 2 & 1\\ 2 & 2 & 1\\ 2 & 3 & 1 \end{matrix},\quad C(\widetilde{T})=\begin{matrix} A & B & C\\ 1 & 2 & 1\\ 1 & 3 & 1\\ 2 & 2 & 1\\ 2 & 3 & 1 \end{matrix},\quad \widetilde{\widetilde{T}}=\begin{matrix} A & B & C\\ 1 & 3 & 1 \end{matrix};\quad \widetilde{\widetilde{T}}\neq T \text{ и}$$

$C(\widetilde{T})\neq C(T)$.

Теорема 2 (правило преобразования разности). Равенство $T_1\cap\widetilde{T}_2=T_1-T_2$ при $T_1-T_2\neq T_\varnothing$ выполняется тогда и только тогда, когда выполняются включения $\forall i\ D_{A_i,T_1}\subseteq D_{A_i,T_2}$.

Доказательство. Необходимость. Пусть выполняется равенство $T_1\cap\widetilde{T}_2=T_1-T_2$. От противного, допустим, что для некоторого q включение $D_{A_q,T_1}\subseteq D_{A_q,T_2}$ не выполняется, то есть существует такой x, что $x\in D_{A_q,T_1}$ и $x\notin D_{A_q,T_2}$. Тогда по определению активного домена $\exists s\in T_1\big|(A_q,x)\in s$. Так как $x\notin D_{A_q,T_2}$, то по определению активного домена $s\notin T_2$, следовательно, $s\in T_1-T_2$. С другой стороны, по определению насыщения $x\notin D_{A_q,T_2}$ влечёт $s\notin C(T_2)$, а поскольку $\widetilde{T}_2=C(T_2)-T_2$, то $s\notin\widetilde{T}_2$, и значит $s\notin T_1\cap\widetilde{T}_2$, что противоречит допущению. Необходимость доказана.

Докажем достаточность утверждения. Пусть выполняются включения $\forall i\, D_{A_i,T_1}\subseteq D_{A_i,T_2}$; докажем, что выполняется равенство $T_1\cap\widetilde{T}_2=T_1-T_2$. В монографии [2] (утверждение 2.2.1 пункт 7) доказано включение $T_1\cap\widetilde{T}_2\subseteq T_1-T_2$, поэтому нам нужно доказать включение $T_1-T_2\subseteq T_1\cap\widetilde{T}_2$. Пусть $s=\{(A_1,d_1),\ldots,(A_k,d_k)\}\in T_1-T_2$. Тогда $s\in T_1$ и $s\notin T_2$. Покажем, что $s\in\widetilde{T}_2$. По определению активного домена выполняются принадлежности $d_1\in D_{A_1,T_1}$, …, $d_k\in D_{A_k,T_1}$. Из условий $\forall i\ D_{A_i,T_1}\subseteq D_{A_i,T_2}$ следует, что $d_1\in D_{A_1,T_2}$, …, $d_k\in D_{A_k,T_2}$. По определению насыщения $s\in C(T_2)$, по определению операции активного дополнения в этом случае включение $s\notin T_2$ влечет $s\in\widetilde{T}_2$, поэтому $s\in T_1\cap\widetilde{T}_2$, то есть равенство выполняется.

Выводы

В работе найдены критерии, при которых в табличных алгебрах выполняются закон двойного отрицания и правило преобразования разности. Результаты работы могут быть использованы в теории обобщенных табличных алгебр и, на наш взгляд, для оптимизации запросов в реляционных базах данных.

Автор благодарит Дмитрия Борисовича Буя за постановку задачи и полезные замечания.

Литература

1. Codd E.F. A Relational Model of Data for Large Shared Data Banks / E.F. Codd // Communications of the ACM. – 1970. – V. 13, №. 6. – P. 377-387.
2. Реляційні бази даних: табличні алгебри та SQL-подібні мови / [В.Н. Редько, Ю.Й. Брона, Д.Б. Буй, С.А. Поляков]. – Київ : Видавничий дім «Академперіодика», 2001. – 198 с.
3. Мейер Д. Теория реляционных баз данных / Д. Мейер ; [пер. с англ.]. – Москва: Мир, 1987. – 608 с.

Literatura

1. Codd E.F. A Relational Model of Data for Large Shared Data Banks / E.F. Codd // Communications of the ACM. – 1970. – V. 13, №. 6. – P. 377-387.
2. Relational data base: table algebras and family of the SQL languages / [V.N. Redko, U.I. Brona, D.B. Buy, S.A. Polyakov]. – Kiev : Publisher «Academperiodica», 2001. – 198 p. (in Ukrainian).
3. D. Maier. The theory of Relational Databases / D. Maier ; [translation from English]. – Moscow : «Mir», 1987. – 608 p. (in Russian).

RESUME

A.S. Senchenko

A Double Complement Law and a Difference Transformation Rule in Table Algebra

In this paper some properties of operations of active complement and difference tables in table algebra is investigated. This algebra is built on the basis of Codd's relational algebra and used for the theoretical ground of query languages of modern tabular databases.

In book [2] is shown that in table algebra the double complement law and the difference transformation rule as inclusions are fulfilled. In this paper the necessary and sufficient conditions are found due to which in table algebra foregoing inclusions are becoming to identities. These conditions are defined in terms of relations between the active domains of attributes of tables.

The results can be used in generalized table algebra theory and for query optimization in relation databases.

Статья поступила в редакцию 05.04.2013.